

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Балашов Александр Геннадьевич

Должность: Проректор по учебной работе

Дата подписания: 02.03.2026 16:25:30

Уникальный программный ключ:

159a23a86061c60c9ea58b6d4823a92e747fed78

Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет

«Московский институт электронной техники»



УТВЕРЖДАЮ

Проректор по учебной работе

А.Г. Балашов

«02» *сентября* 2026 г.

**ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА -
ПРОГРАММА ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ
«КОМПЬЮТЕРНАЯ ЛИНВИСТИКА (NLP)»**

Объем 438 часов

Москва 2026 г.

1. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

1.1. Цель реализации программы

Цель программы – формирование у слушателей профессиональных компетенций, необходимых для выполнения нового вида профессиональной деятельности и приобретения новой квалификации.

1.2. Характеристика новой квалификации / нового вида профессиональной деятельности

Выпускникам программы присваивается квалификация «Программист».

Выпускник программы будет готов к выполнению трудовой деятельности в области создания прикладных информационных систем с использованием методов машинного обучения и интеллектуальной обработки данных в качестве специалиста по обработке естественного языка (инженера по обработке естественного языка), и выполнению трудовых функций в области разработки моделей машинного обучения для обработки естественного языка и адаптации моделей машинного обучения для разработки прикладных решений с использованием средств обработки естественного языка.

Перечень и характеристика компетенций, подлежащих совершенствованию, и (или) перечень новых компетенций, формирующихся в результате освоения программы:

Область профессиональной деятельности: 06 Связь, информационные и коммуникационные технологии.

Наименование вида экономической деятельности: 62.01 Разработка компьютерного программного обеспечения.

Укрупненная группа специальностей: 09.00.00 «Информатика и вычислительная техника».

1.3. Требования к результатам освоения программы

Дополнительная профессиональная программа (программа профессиональной переподготовки) «Компьютерная лингвистика (NLP)» разработана в соответствии с:

- федеральным государственным образовательным стандартом высшего образования - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника, утвержденным приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. № 929 (далее – ФГОС ВО);

- профессиональным стандартом 06.001 "Программист", утвержденным приказом Министерства труда и социальной защиты Российской Федерации от 20 июля 2022 г. № 424н (зарегистрирован Министерством юстиции Российской Федерации 22 августа 2022 г., регистрационный N 69720).

Результаты освоения программы профессиональной переподготовки представлены в Таблице 1.

Таблица 1 - Результаты освоения программы профессиональной переподготовки

Код и формулировка профессиональной /профессионально-специализированной компетенции	Трудовая функция в соответствии с профессиональным стандартом		Индикаторы достижения компетенций
	Наименование	Код	
ПК-1 Применяет языки программирования и настраиваемые программные инструменты для автоматизации процессов в профессиональной деятельности	Проектирование компьютерного программного обеспечения	D/03.6	Знания: понятий и основных технических приемов матричной алгебры, аналитической геометрии, булевой алгебры, теории линейных операторов и квадратичных форм; элементов теории кодирования; синтаксиса Python и основных конструкций; методов работы с файлами и функциями; принципов объектно-ориентированного программирования; методов обработки ошибок; основ многопоточности и сетевого взаимодействия. Умения: применять современные методы компьютерной реализации геометрических и

Код и формулировка профессиональной /профессионально-специализированной компетенции	Трудовая функция в соответствии с профессиональным стандартом		Индикаторы достижения компетенций
	Наименование	Код	
			алгебраических моделей к решению задач; создавать отчеты о проделанной работе с пояснениями, графиками и выводами; писать программы на Python для анализа данных; использовать библиотеки Python для обработки и анализа информации; проектировать и реализовывать классы и модули; отлаживать и тестировать код; применять многопоточность и создавать сетевые приложения. <i>Опыт деятельности:</i> использования синтаксиса и библиотек Python/SQL/других языков в соответствии с задачей; интеграции отдельных программных модулей в единое приложение, оформление кода с комментариями.
ПК-2 Разрабатывает отдельные части проектов по применению искусственного интеллекта и машинного обучения	Проектирование компьютерного программного обеспечения	D/03.6	<i>Знания:</i> современных методов обработки естественного языка; алгоритмов построения векторных представлений текста; методов предобработки текстовых данных; принципов работы нейронных сетей для обработки естественного языка; основ работы с большими языковыми моделями; основных понятий и теорий фонетики и фонологии, включая глубинные представления, генеративную фонологию и теорию оптимальности; строения морфологической системы языка и методы морфологического анализа; принципов генеративной и X-штрих грамматики; семантических типов, рестрикторов и кванторов и их роль в структуре предложения; основных подходов в обработке естественного языка и вероятностных моделей. <i>Умения:</i> применять библиотеки для обработки текстов и работы с нейронными сетями; разрабатывать и оптимизировать пайплайны для решения задач в сфере обработки естественного языка; работать с API больших языковых моделей; разрабатывать программные модули для обработки естественного языка; проводить фонетический и морфологический анализ слов и предложений; строить деревья зависимостей и применять методы генеративной грамматики; использовать библиотеки для NLP для предобработки текста и построения моделей классификации и регрессии; разрабатывать и оценивать простые языковые модели; анализировать результаты машинного обучения, корректно интерпретировать метрики. <i>Опыт деятельности:</i> выбора готовой модели ИНС (искусственной нейронной сети) или разработки собственной, обработки данных для обучения/тестирования модели, оценка качества работы модели по метрикам.
ПК-3 Применяет программные алгоритмы для обработки данных, подбирает и	Проектирование компьютерного программного обеспечения	D/03.6	<i>Знания:</i> основных этапов обработки текстов; методов распознавания именованных сущностей и оценки качества; методов морфологического анализа и POS-тегирования, различий между

Код и формулировка профессиональной /профессионально-специализированной компетенции	Трудовая функция в соответствии с профессиональным стандартом		Индикаторы достижения компетенций
	Наименование	Код	
использует специализированные библиотеки.			стеммингом и лемматизацией; основ конечных автоматов (FSA/FST) для морфологических задач; принципов работы с API и сбора данных из открытых источников; контекстно-свободных и зависимых грамматик, алгоритмов CYK и Earley, методов извлечения коллокаций; тематического моделирования (LDA) и мер сходства текстов; методов классификации, суммаризации и анализа тональности текстов; особенностей обработки многоязычных текстов и алгоритмов извлечения ключевых слов; основ анализа текстов социальных медиа и мер сходства текстов; основных принципов функционализма; теории фреймов Ч. Филлмора и её применение в лингвистике; основных моделей грамматики конструкций (Филлмора, Лакоффа-Гольдберга, Крофта); технологии работы с лингвистическими корпусами; корпусных методов исследований; основных лингвистических ресурсов и баз данных (FrameNet, НКРЯ). <i>Умения:</i> выполнять предобработку текста; реализовывать методы rule-based и статистического NER, POS-тегирования и стемминга/лемматизации на Python; строить конечные автоматы для морфологических задач; получать данные через API и web-скрейпинг, соблюдать этические нормы; реализовывать простейшие парсеры на базе CFG и алгоритмы переходного-ориентированного анализа; применять методы тематического моделирования и вычислять меры сходства; строить классификаторы и проводить анализ тональности; реализовывать экстрактивные алгоритмы суммаризации и простые QA-системы; работать с многоязычными данными и извлекать ключевые слова; выполнять сетевой анализ текстов и оценивать сходство текстов; применять принципы функциональной и когнитивной лингвистики при анализе языкового материала; использовать корпусные методы исследования для решения поставленных лингвистических задач; работать с лингвистическими ресурсами (FrameNet, НКРЯ) для извлечения и систематизации данных; разрабатывать простые аналитические алгоритмы для обработки языковых данных с использованием стандартных библиотек; применять модели грамматики конструкций и фреймовой семантики для интерпретации и описания языковых явлений. <i>Опыт деятельности:</i> анализа входных данных и выбора алгоритма их обработки, реализации алгоритма в виде работающего кода, проведения тестирования и оптимизации решения.
ПК-4 Использует СУБД при разработке ПО.	Проектирование компьютерного	D/03.6	<i>Знания:</i> классификации и архитектуры СУБД; основных моделей данных и модель ANSI/SPARC;

Код и формулировка профессиональной /профессионально-специализированной компетенции	Трудовая функция в соответствии с профессиональным стандартом		Индикаторы достижения компетенций
	Наименование	Код	
	программного обеспечения		основ реляционной алгебры; синтаксиса SQL и его основных конструкций; принципов работы NoSQL баз данных. <i>Умения:</i> проектировать и создавать структуры баз данных; формировать SQL-запросы различной сложности; выполнять операции выборки, обновления и удаления данных; интегрировать СУБД с клиентскими приложениями; использовать NoSQL базы данных для решения прикладных задач. <i>Опыт деятельности:</i> разработки и оптимизации запросов; работы с реляционными и NoSQL СУБД; взаимодействия СУБД с внешними приложениями; проектирования структур данных.

1.4. Требования к уровню подготовки поступающего на обучение, необходимому для освоения программы

К обучению по программе допускаются лица, имеющие среднее профессиональное или высшее образование или получающие среднее профессиональное или высшее образование (при наличии соответствующей справки с указанием года окончания).

1.5. Трудоемкость обучения

Нормативная трудоемкость обучения по данной программе – 438 часов, включая все виды аудиторной и внеаудиторной (самостоятельной) учебной работы слушателя.

1.6. Форма обучения

Форма обучения: очная, очно-заочная, с использованием электронного обучения и дистанционных образовательных технологий.

1.7. Режим занятий

Обучение осуществляется без отрыва от работы. Учебная нагрузка устанавливается не более 12 часов в неделю, включая все виды аудиторной и внеаудиторной (самостоятельной) учебной работы слушателя. Продолжительность одного академического часа занятий — 40 минут.

2. СОДЕРЖАНИЕ ПРОГРАММЫ

2.1. Учебный план программы

№ п/п	Наименование учебных дисциплин / модулей	Всего, час	В том числе						Форма аттестации (час.)
			аудиторных				самостоятельная работа		
			лекции		Практические / лабораторные занятия				
					очная	очно- заочная	очная	очно- заочная	
1.	Высшая математика на Python	42	16	8	16	16	8	16	Зачет (2)
2.	Python для анализа данных	66	16	8	32	32	8	24	Зачет (2)
3.	Структурная лингвистика	66	16	8	32	32	18	24	Зачет (2)
4.	Функциональные и когнитивные модели в лингвистике	70	16	8	32	32	22	28	Зачет (2)
5.	SQL и NoSQL базы данных	50	8	4	20	20	22	24	Зачет (2)
6.	Автоматическая обработка текста	72	14	8	28	28	30	34	Зачет (2)
7.	Нейронные сети для обработки естественного языка	66	16	8	32	32	18	24	Зачет (2)
8.	Итоговая аттестация (демонстрационный экзамен)	6							
9.	Всего	438	102	52	192	192	126	174	14

2.2. Календарный учебный график

Учебные недели/дни																																						
Наименование учебных дисциплин (модулей), практик		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Высшая математика на Python		А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	ПА																			
Python для анализа данных		А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	ПА																			
Структурная лингвистика		А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	ПА																			
Функциональные е и когнитивные модели в лингвистике		А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	ПА																			
SQL и NoSQL базы данных																				А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	ПА	
Автоматическая обработка текста																				А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	ПА	
Нейронные сети для обработки естественного языка																				А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	А	ПА	
Итоговая аттестация																																						ИА

Условные обозначения:

А – аудиторное занятие (лекция, практическое занятие, самостоятельная работа); П – практика; С – стажировка; ПА – промежуточная аттестация; ИА – итоговая аттестация

3. РАБОЧИЕ ПРОГРАММЫ УЧЕБНЫХ ДИСЦИПЛИН (МОДУЛЕЙ), ПРАКТИК/ СТАЖИРОВОК

3.1. Рабочая программа дисциплины «Высшая математика на языке Python»

Рабочая программа модуля «Высшая математика на языке Python» направлена на формирование студентами знаний и умений в области решения профессиональных задач на Python.

Тема 1 «Элементы линейной алгебры»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
1.1.	Элементы линейной алгебры. Линейное пространство двоичных векторов.	2	ЭО (2 часа)
1.2.	Матрицы. Операции над матрицами. Обратная матрица. Решение систем линейных уравнений.	2	ЭО (2 часа)
1.3.	Поле Z_2 . Решение систем линейных уравнений над полем Z_2 .	2	ЭО (2 часа)
Перечень лабораторных работ			
1.1.	Построение графиков. Визуализация векторов. Работа с векторами в Python.	2	ДОТ (2 часа)
1.2.	Представление и вывод матриц в Python. Работа с матрицами в Python. Операции над матрицами в Python.	2	ДОТ (2 часа)
1.3.	Составление и решение систем линейных уравнений в Python. Работа с полем Z_2 в Python.	2	ДОТ (2 часа)
Тема 2 «Алгебра логики»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
2.1.	Алгебра логики. Высказывания и логические операции над ними. Функции алгебры логики. Формулы алгебры логики.	2	ЭО (2 часа)
2.2.	СДНФ и СКНФ. Равносильные преобразования формул. бинарные отношения на множестве. Свойства бинарных отношений.	2	ЭО (2 часа)
2.3.	Отношения эквивалентности и порядка. Частично упорядоченные множества.	2	ЭО (2 часа)
Перечень лабораторных работ			
2.1.	Логические операторы в Python. Составление и вывод таблиц истинности в Python. Модули библиотеки sympy для решения задач булевой алгебры.	2	ДОТ (2 часа)
2.2.	Составление и анализ СДНФ, СКНФ и полинома Жегалкина в Python. Работа со множествами и бинарными отношениями в Python.	2	ДОТ (2 часа)
2.3.	Иллюстрация бинарных отношений с помощью матриц, графиков и орграфов в Python.	2	ДОТ (2 часа)
Тема 3 «Элементы теории кодирования»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов

Содержание лекционных занятий			
3.1.	Элементы теории кодирования. Линейные коды.	2	ЭО (2 часа)
3.2.	Алгоритм кодирования Шеннона – Фано. Коды Хэмминга.	2	ЭО (2 часа)
Перечень лабораторных работ			
3.1.	Кодирование текстовых данных и изображений в Python. Алгоритм кодирования Шеннона – Фано в Python.	2	ДОТ (2 часа)
3.2.	Алгоритм кодирования Хэмминга в Python.	2	ДОТ (2 часа)

3.1.1. Учебно-методическое и информационное обеспечение учебной дисциплины

Для теоретической подготовки используются учебная литература и комплект теоретических материалов, в виде электронных презентаций к лекциям, содержащие основные определения, математический аппарат и иллюстративный материал.

Для подготовки к лабораторным работам используется лабораторный практикум, содержащий описание методов и подходов к выполнению лабораторных заданий.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

Перечень учебной литературы:

1) Бугров Я.С. Высшая математика: В 3-х т.: Учеб. для вузов. Т. 1: Дифференциальное и интегральное исчисление: в 2 кн.: Кн. 1 / Я.С. Бугров, С.М. Никольский. — 7-е изд., стер. — М.: Юрайт, 2020. — 253 с.

2) Ильин В.А. Линейная алгебра: Учебник для вузов / Ильин В.А., Позняк Э.Г. — 6-е изд., стер. М.: Физматлит, 2010. — 278 с. — (Курс высшей математики и математической физики. Вып. 4)

3) Сборник задач по математике для втузов. В 4-х частях: Учебное пособие для втузов. /Под общ. ред. А.В. Ефимова и А.С. Поспелова. — 5-е изд. испр. — М.: Физматлит, 2009. — Ч. 1.

4) Бронштейн И.Н. Справочник по математике для инженеров и учащихся вузов: Учеб. пособие / И.Н. Бронштейн, К.А. Семендяев. — СПб.: Лань, 2010. — 608 с.

Информационные ресурсы:

1) Интерактивная среда для выполнения Python-кода с возможностью визуализации и документирования. URL: <https://jupyter.org> (дата обращения: 27.07.2025).

2) Официальная документация Python 3. URL: <https://docs.python.org/3> (дата обращения: 27.07.2025).

3.1.2. Материально-техническое обеспечение учебной дисциплины

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE). Возможен формат BYOD.

3.1.3. Система контроля и оценивания

Оценка качества освоения дисциплины включает промежуточную и итоговую аттестацию обучающихся.

Промежуточный контроль знаний проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся.

Итоговая аттестация по дисциплине проводится в форме тестирования, расположенного на СДО iSpring по адресу <https://edu.miet.ru>. Для получения зачета по дисциплине необходимо верно ответить не менее, чем на 70% вопросов итогового тестирования.

Пример типовых тестовых заданий:

№ п/п	Вопрос	Варианты ответа
1	Что такое модуль вектора в декартовой системе координат?	1. длина вектора 2. угол поворота вектора относительно оси Oх

		3. сумма координат вектора 4. сумма модулей координат вектора
2	Что из нижеперечисленного является логическим высказыванием?	1. «Чему равен определитель матрицы?» 2. «Определитель матрицы равен 0» 3. «Обратная матрица может существовать только у невырожденных матриц» 4. «Модуль матрицы» 5. «Матрица логических высказываний»
3	Определите по матрице свойства бинарного отношения $M = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$	1) Транзитивность 2) Рефлексивность 3) Симметричность 4) Антирефлексивность 5) Антисимметричность 6) Полнота

Пример типового практического задания:

Создайте функцию, реализующую алгоритм приведения формулы, заданной списком ее значений на всех наборах переменных, к СКНФ. Проверьте работу функции с помощью библиотечной функции `sympy.logic.boolalg.is_cnf()`.

Данное практическое задание предполагает выполнение следующих этапов:

- 1) Задать формулу списком ее значений на всех наборах переменных.
- 2) Реализовать алгоритм конструирования элементарных дизъюнкций на основе набора переменных.
- 3) Реализовать алгоритм конструирования Совершенной Конъюнктивной Нормальной Формы из элементарных дизъюнкций на основе списка значений данной формулы.
- 4) Проверить работу реализованных функции с помощью библиотечной функции `sympy.logic.boolalg.is_cnf()`.
- 5) Оформить и защитить отчет по практическому заданию.

3.2. Рабочая программа дисциплины

«Python для анализа данных»

Рабочая программа модуля «Python для анализа данных» нацелена на формирование у обучающихся знаний и навыков программирования на языке Python для решения прикладных задач.

Тема 1 «Введение в Python»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
1.1.	История программирования, сферы применения, особенности синтаксиса, установка и настройка среды (PyCharm, VS Code, Jupyter Notebook)	2	ЭО (2 часа)
Перечень лабораторных работ			
1.1.	Настройка окружения, «Hello, world!», работа в интерактивном режиме	4	ЭО (2 часа), ДОТ (2 часа)
Тема 2 «Ввод и вывод данных»		Кол-во часов	Образовательные технологии, в

			том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
2.1.	Методы input(), print(), форматирование строк, escape-последовательности	2	ЭО (2 часа)
Перечень лабораторных работ			
2.1.	Программа ввода/вывода с форматированием, создать программу, принимающую имя и возраст и выводящую приветствие	4	ЭО (2 часа), ДОТ (2 часа)
Тема 3 «Работа с файлами»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
3.1.	Чтение и запись файлов, режимы, with open	2	ЭО (2 часа)
Перечень лабораторных работ			
3.1.	Программа чтения данных и записи результата в новый файл, прочитать CSV, извлечь нужные столбцы, записать в отдельный файл	4	ЭО (2 часа), ДОТ (2 часа)
Тема 4 «Работа с функциями»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
4.1.	Определение функций, параметры, возвращаемые значения, аннотации типов	2	ЭО (2 часа)
Перечень лабораторных работ			
4.1.	Функции для обработки данных (сумма, среднее, фильтрация), создать библиотеку функций для анализа текста	4	ЭО (2 часа), ДОТ (2 часа)
Тема 5 «ООП в Python»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
5.1.	Классы, объекты, наследование, полиморфизм, инкапсуляция, магические методы	2	ЭО (2 часа)
Перечень лабораторных работ			
5.1.	Класс для работы с данными (загрузка, фильтрация, сортировка), создать два класса, моделирующие предметную область.	4	ЭО (2 часа), ДОТ (2 часа)
Тема 6 «Анализ ошибок»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
6.1.	Исключения, try/except/finally, собственные исключения	2	ЭО (2 часа)
Перечень лабораторных работ			
6.1.	Программа с обработкой ошибок при чтении файла, модифицировать программу для корректной работы при ошибках ввода	4	ЭО (2 часа), ДОТ (2 часа)
Тема 7 «Прикладные задачи»		Кол-во часов	Образовательные технологии, в

			том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
7.1.	Библиотеки pandas, numpy	2	ЭО (2 часа)
Перечень лабораторных работ			
7.1.	Анализ данных из CSV и генерация отчета, решение задач анализа данных с pandas и matplotlib	4	ЭО (2 часа), ДОТ (2 часа)
Тема 8 «Многопоточность и сетевые приложения»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
8.1.	Обзор: threading, concurrent.futures, socket, клиент-сервер, простые протоколы	2	ЭО (2 часа)
Перечень лабораторных работ			
8.1.	Добавить многопоточность в существующую программу.	4	ЭО (2 часа), ДОТ (2 часа)
Промежуточная аттестация		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
9.1.	Промежуточная аттестация в формате тестирования	2	ЭО (2 часа)

3.2.1. Учебно-методическое и информационное обеспечение учебной дисциплины

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

Перечень учебной литературы:

- 1) Сузи, Р. А. Язык программирования Python: учебное пособие / Р. А. Сузи. — 2-е изд. — Москва: ИНТУИТ, 2016. — 350 с. — URL: <https://e.lanbook.com/book/100546> (дата обращения: 19.07.2025). - Режим доступа: для авториз. пользователей МИЭТ.
- 2) Лутц М. Изучаем Python. — 5-е изд. — СПб.: Питер, 2022. — 1600 с.
- 3) Седжвик Р., Уэйн К. Алгоритмы на Python. — СПб.: Питер, 2021. — 928 с.
- 4) Beazley D. Python Cookbook. — 4th ed. — O'Reilly Media, 2023. — 706 p.
- 5) Downey A. Think Python: How to Think Like a Computer Scientist. — 3rd ed. — Green Tea Press, 2023. — 300 p.

Информационные ресурсы:

- 1) Интерактивная среда для выполнения Python-кода с возможностью визуализации и документирования. URL: <https://jupyter.org> (дата обращения: 27.07.2025).
- 2) Официальная документация Python 3. URL: <https://docs.python.org/3> (дата обращения: 27.07.2025).

3.2.2. Материально-техническое обеспечение учебной дисциплины

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE). Возможен формат BYOD.

3.2.3. Система контроля и оценивания

Оценка качества освоения дисциплины включает промежуточную и итоговую аттестацию обучающихся.

Промежуточный контроль знаний проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся.

Итоговая аттестация по дисциплине проводится в форме тестирования, расположенного на СДО iSpring по адресу <https://edu.miet.ru>. Для получения зачета по дисциплине необходимо верно ответить не менее, чем на 70% вопросов итогового тестирования.

Пример типовых тестовых заданий:

№ п/п	Вопрос	Варианты ответа
1	Какой тип данных в Python изменяемый?	1. tuple 2. str 3. list 4. int
2	Что из нижеперечисленного является логическим высказыванием?	1. cut() 2. trim() 3. strip() 4. remove()
3	Какой модуль используется для многопоточности?	1. multiprocessing 2. threading 3. asyncio 4. parallel
4	Что такое инкапсуляция в ООП?	1. Запрет наследования 2. Объединение данных и методов в одном классе 3. Автоматическое создание методов 4. Копирование объектов
5	Как в Python обрабатываются исключения?	1. try/finally 2. try/catch 3. try/except 4. error handling block

Примеры типовых практических заданий:

- 1) Напишите функцию, которая принимает список чисел и возвращает их среднее значение.
- 2) Реализуйте класс Employee с методами добавления и отображения информации о сотруднике.
- 3) Разработайте скрипт, который скачивает данные по API и сохраняет их в файл.
- 4) Напишите многопоточную программу, которая обрабатывает несколько файлов одновременно.
- 5) Создайте простое клиент-серверное приложение для обмена сообщениями.

3.3. Рабочая программа дисциплины

«Структурная лингвистика»

Рабочая программа модуля «Структурная лингвистика» нацелена на изучение звуковой, морфологической и синтаксической структуры языка, рассматриваются основные фонологические теории, основы морфологии и генеративной грамматики, семантические типы и отношения, а также методы обработки языковых структур с помощью статистических и машинных моделей. Целью программы является развитие у студентов понимания теоретических моделей и формирование навыков применения алгоритмов для анализа реальных лингвистических данных.

Тема 1 «Фонетика и фонология»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
1.1.	Фонетика; глубинные представления; генеративная фонология; теория оптимальности.	2	ЭО (2 часа)
Перечень лабораторных работ			
1.1.	Введение в курс; обзор библиотек для НЛП; знакомство с IPA и Unicode; простейшая генерация транскрипции	4	ЭО (2 часа), ДОТ (2 часа)
Тема 2 «Фонетика и морфология»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
2.1.	Фонетика; теория оптимальности	2	ЭО (2 часа)
Перечень лабораторных работ			
2.1.	Морфологический анализ; деревья зависимостей, выполнение морфологического анализа слов с использованием существующих анализаторов; построение деревьев зависимостей для простых предложений.	4	ЭО (2 часа), ДОТ (2 часа)
Тема 3 «Грамматика и семантика»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
3.1.	Грамматика; генеративная грамматика	2	ЭО (2 часа)
Перечень лабораторных работ			
3.1.	Семантика; логика и векторные представления, построение синтаксических деревьев; реализация простых моделей векторных представлений и вычисление семантической близости.	4	ЭО (2 часа), ДОТ (2 часа)
Тема 4 «Синтаксис и предобработка»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
4.1.	Определение функций, параметры, возвращаемые значения, аннотации типов	2	ЭО (2 часа)
Перечень лабораторных работ			
4.1.	Предобработка текста; основные этапы машинного обучения, подготовка текстового корпуса: токенизация, нормализация, удаление стоп-слов; исследование влияния предобработки на качество анализа.	4	ЭО (2 часа), ДОТ (2 часа)
Тема 5 «Семантические типы»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
5.1.	Семантика; семантические типы	2	ЭО (2 часа)
Перечень лабораторных работ			
5.1.	Классификация; применение простых классификаторов.	4	ЭО (2 часа), ДОТ

	Создание датасета и реализация модели классификации для лингвистических задач; анализ метрики точности и полноты.		(2 часа)
Тема 6 «Кванторы и регрессия»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
6.1.	Семантика; рестрикторы, кванторы, сферы применения	2	ЭО (2 часа)
Перечень лабораторных работ			
6.1.	Регрессия; построение регрессионных моделей. применение регрессионной модели для оценки параметров в текстовых данных; сравнение результатов разных моделей.	4	ЭО (2 часа), ДОТ (2 часа)
Тема 7 «Подходы в NLP (natural language processing, НЛП)»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
7.1.	НЛП; подходы в НЛП; вероятностные модели	2	ЭО (2 часа)
Перечень лабораторных работ			
7.1.	Языковое моделирование; введение в n-gram модели, реализация простой n-gram языковой модели; анализ частотных распределений и энтропии.	4	ЭО (2 часа), ДОТ (2 часа)
Тема 8 «Разведочный анализ»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
8.1.	НЛП; введение в машинное обучение, разведочный анализ и обработку признаков	2	ЭО (2 часа)
Перечень лабораторных работ			
8.1.	Мини-проект: выбор датасета, выполнение разведочного анализа, предобработка, выбор алгоритма машинного обучения, оценка метрик	4	ЭО (2 часа), ДОТ (2 часа)
Промежуточная аттестация		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
9.1.	Промежуточная аттестация в формате тестирования	2	ЭО (2 часа)

3.3.1. Учебно-методическое и информационное обеспечение учебной дисциплины

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

Перечень учебной литературы:

- 1) Керстен Гробок, Катрин Хандке. Введение в структурную лингвистику. — М.: Языки славянской культуры, 2010.
- 2) Н. Хомский. Аспекты теории синтаксиса. — М.: Прогресс, 1972.
- 3) П. Адлер. Математическая лингвистика. — М.: Едиториал УРСС, 2005.

4) Daniel Jurafsky, James H. Martin. Speech and Language Processing. 3rd ed. — Prentice Hall, 2023.

5) Christopher Manning, Hinrich Schütze. Foundations of Statistical Natural Language Processing. — MIT Press, 1999.

3.3.2. Материально-техническое обеспечение учебной дисциплины

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE). Предусмотрен формат BYOD.

3.3.3. Система контроля и оценивания

Оценка качества освоения дисциплины включает промежуточную и итоговую аттестацию обучающихся.

Промежуточный контроль знаний проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся.

Итоговая аттестация по дисциплине проводится в форме тестирования, расположенного на СДО iSpring по адресу <https://edu.miet.ru>. Для получения зачета по дисциплине необходимо верно ответить не менее, чем на 70% вопросов итогового тестирования.

Пример типовых тестовых заданий:

№ п/п	Вопрос	Варианты ответа
1	Что является основной единицей фонологического анализа в структурной лингвистике?	a) Морфема b) Фонема c) Аллофон d) Графема
2	Какой метод используется для выявления минимальных смыслоразличительных единиц языка?	a) Метод трансформаций b) Метод компонентного анализа c) Метод оппозиции d) Метод статистического анализа
3	Кто считается основателем Пражской лингвистической школы?	a) Ноам Хомский b) Роман Jakobson c) Фердинанд де Соссюр d) Леонард Блумфилд
4	Что в структурной лингвистике понимается под дистрибуцией?	a) Распределение языковых единиц в тексте b) Соотношение звуков в фонологической системе c) Набор морфологических парадигм d) Порядок слов в предложении
5	Как называется процесс выделения синтагм в потоке речи?	a) Сегментация b) Ассимиляция c) Деривация d) Редукция

Примеры типовых практических заданий:

1) Проанализируйте приведённый отрывок текста, выделите фонемы и укажите их оппозиции.

- 2) Постройте морфологическую схему слова «непрочитанный», обозначив корень, приставку, суффикс и окончание.
- 3) Произведите синтаксический разбор предложения «Студенты анализируют текст с точки зрения его структуры».
- 4) Используя метод компонентного анализа, опишите семантические различия между словами «дом» и «здание».
- 5) Разработайте схему структурных уровней языка, от фонетического до прагматического, и кратко опишите каждый уровень.

3.4. Рабочая программа дисциплины

«Функциональные и когнитивные модели в лингвистике»

Рабочая программа модуля «Функциональные и когнитивные модели в лингвистике» направлена на получение студентами знаний и умений в области функционального и когнитивного анализа языка, включая изучение принципов функционализма, системно-функциональной лингвистики, фреймовой семантики, когнитивной грамматики и корпусных методов исследования.

Тема 1 «Основные принципы функционализма»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
1.1.	Принципы функционализма: объяснение языковой формы через функции, типологическая ориентированность, эмпиризм, количественные методы, междисциплинарность, иконизм, экономия, мотивация грамматики дискурсом. Проблема грамматикализации.	2	ДОТ (2 часа)
Перечень лабораторных работ			
1.1.	Исследование экономии и иконичности в языке на основе корпусных данных. Выявление примеров проявления экономии и иконичности количественными методами в текстах различных стилей, жанров и эпох. Исследование явления экономии: подсчет частотности сокращений и использования клише, сравнение длины слов и предложений в текстах разных стилей, подсчет статистики использования эллипсиса. Исследование явления иконичности: выявление лексических и морфологических повторов, а также закономерностей их использования в текстах различных стилей; исследование порядка слова и их соответствия последовательности передаваемых событий; изучение явления звукоподобия.	2	ЭО (2 часа)
1.2.	Исследование принципа грамматикализации (перехода лексических единиц в грамматические) на основе корпусных данных. Частотный анализ: сравнение частотности слова в исходном и грамматикализованном значении. Анализ проявления грамматикализации в диахронии: контекстуальный анализ употребления слова в разных временных срезах. Поиск в корпусе примеров грамматикализации разных типов: фонетическая редукция, дисемантизация, расширение контекстов употребления, фиксация синтаксической позиции.	2	ЭО (2 часа)
Тема 2 «Системно-функциональная лингвистика (СФЛ)»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов

Содержание лекционных занятий			
2.1.	Модель языка в СФЛ. Метафункции: понятийная, межличностная, текстуальная. Кардиффская модель функциональной грамматики (М. Халлидей и др.), лексико-грамматика, классы и функции, грамматикализация.	2	ДОТ (2 часа)
Перечень лабораторных работ			
2.1.	Анализ текста с точки зрения метафункций СФЛ.	2	ЭО (2 часа)
2.2.	Применение кардиффской модели для анализа предложений.	2	ЭО (2 часа)
Тема 3 «Функциональная грамматика»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
3.1.	Функциональная грамматика на уровне предложения: предложение как сообщение; информационная структура предложения; предложение как диалог; предложение как репрезентация. Функциональная грамматика на уровне дискурса.	2	ДОТ (2 часа)
Перечень лабораторных работ			
3.1.	Разбор информационной структуры текста.	2	ЭО (2 часа)
3.2.	Анализ дискурса с позиций функциональной грамматики. Выявление средств связности в тексте, анализ роли грамматических структур в организации дискурса.	2	ЭО (2 часа)
Тема 4 «Фреймовая семантика и когнитивные модели»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
4.1.	Понятие фрейма. Фреймовая семантика Ч. Филлмора. Структура фрейма. Сцены и скрипты. Фреймы в дискурсе. Фреймовый анализ с применением корпусов. Фреймовый подход к анализу терминологии (П. Фабер и др.).	2	ДОТ (2 часа)
Перечень лабораторных работ			
4.1.	Фреймовый анализ терминологии с использованием корпусов.	2	ЭО (2 часа)
4.2.	Построение схемы предметной онтологии на основе фреймов.	2	ЭО (2 часа)
Тема 5 «Идеализированные когнитивные модели (ИКМ)»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
5.1.	Кластерные и метонимические ИКМ. Лингвокультурная специфика ИКМ. Когнитивный подход к лексической семантике: полисемия, гиперо- и гипонимия, антонимия. Проект FrameNet и аналогичные проекты. Опыт применения фреймового анализа для построения предметных онтологий.	2	ДОТ (2 часа)
Перечень лабораторных работ			
5.1.	Сбор данных для построения формальной онтологии.	2	ЭО (2 часа)
5.2.	Создание скрипта для разработки формальной онтологии.	2	ЭО (2 часа)
Тема 6 «Когнитивная грамматика»		Кол-во часов	Образовательные технологии, в

			том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
6.1.	Субъективистский подход к значению. Профиль - база, траектор - ориентир, когнитивная область. Когнитивные точки отсчета. Метонимия. Активная зона. Когнитивный подход к исследованию грамматических явлений: время, вид, залог, модальность с точки зрения когнитивной лингвистики.	2	ДОТ (2 часа)
Перечень лабораторных работ			
6.1.	Анализ концептов с точки зрения понятий когнитивной грамматики: определение терминов «профиль», «база», «траектор», «ориентир» с использованием языковых примеров и схем.	2	ЭО (2 часа)
6.2.	Анализ грамматических категорий (время, вид, залог) в заданном тексте с когнитивной точки зрения.	2	ЭО (2 часа)
Тема 7 « Грамматика конструкций »		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
7.1.	Модели Филлмора, Лакоффа-Гольдберга, У. Крофта. Коллострукционный анализ. Грамматика конструкции и понятие коллострукции. Коллострукционный анализ с применением корпусных данных. Речевой опыт как основа формирования и функционирования грамматических категорий.	2	ДОТ (2 часа)
Перечень лабораторных работ			
7.1.	Применение модели Филлмора для анализа конструкций в заданном тексте.	2	ЭО (2 часа)
7.2.	Корпусный анализ коллокаций и коллострукций.	2	ЭО (2 часа)
Тема 8 «Корпусные методы в когнитивной лингвистике»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
8.1.	Корпусный подход в когнитивной лингвистике. Анализ статистических данных в когнитивной лингвистике. Эффект частотности. Корпусный анализ коллокаций в когнитивной лингвистике. Коллексемный анализ. Инструменты для корпусных исследований.	2	ДОТ (2 часа)
Перечень лабораторных работ			
8.1.	Использование корпусных инструментов (AntConc, Sketch Engine и др.) для анализа данных. Сравнение существующих инструментов для анализа корпусов с возможностями инструментов из области анализа данных.	2	ЭО (2 часа)
8.2.	Сравнительный анализ частотности языковых явлений в разных корпусах: выбор явления и составление сравнительно-сопоставительных таблиц и графиков.	2	ЭО (2 часа)
Промежуточная аттестация		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
9.1.	Промежуточная аттестация в формате тестирования	2	ЭО (2 часа)

3.4.1. Учебно-методическое и информационное обеспечение учебной дисциплины

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

Перечень учебной литературы:

- 1) Halliday M.A.K. An Introduction to Functional Grammar. – Oxford University Press Inc, 2004.
URL:
2) https://www.uel.br/projetos/ppcat/pages/arquivos/RECURSOS/2004_HALLIDAY_MATTHIESSEN_An_Introduction_to_Functional_Grammar.pdf (дата обращения 04.08.2025)
- 3) Fillmore C. J. Frame Semantics. – Fillmore Department of Linguistics University of Californial Berkeley, 1976. URL: https://brenocon.com/Fillmore%201982_2up.pdf (дата обращения 04.08.2025)
- 4) Лакофф Д. Женщины, огонь и опасные вещи: Что категории языка говорят нам о мышлении / Пер. с англ. И. Б. Шатуновского. — М.: Языки славянской культуры, 2004. — 792 с. — (Язык. Семиотика. Культура). URL: https://psv4.userapi.com/s/v1/d/GqJZgMIGfm-5ba0Vqe-42RUHQs1tTXLxGPGTMUsIkp7HlImpjMXzymfD8ubXAqPIU03SJSbxAkz9hd_PZmvpOOZhFmUDv97uH0Dftti7oPnRylRd/2004_Lakoff_Dzh_Zhenschiny_ogon_i_opasnye_veschi_Chto_kategorii_yazyka_govorya_t_nam_o_myshlenii.pdf (дата обращения 04.08.2025)
- 5) Кибрик А.А. Когнитивные исследования по дискурсу. — М.: 1994. — № 5. — С. 126-139.
URL: <https://vja.ruslang.ru/ru/archive/1994-5/126-139> дата обращения 04.08.2025)

Информационные ресурсы:

- 1) Cognitive linguistics. URL: <https://www.cognitivelinguistics.org/en/about-cognitive-linguistics> (дата обращения: 04.08.2025).

3.4.2. Материально-техническое обеспечение учебной дисциплины

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE). Предусмотрен формат BYOD.

3.4.3. Система контроля и оценивания

Оценка качества освоения дисциплины включает промежуточную и итоговую аттестацию обучающихся.

Промежуточный контроль знаний проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся.

Итоговая аттестация по дисциплине проводится в форме тестирования, расположенного на СДО iSpring по адресу <https://edu.miet.ru>. Для получения зачета по дисциплине необходимо верно ответить не менее, чем на 70% вопросов итогового тестирования.

Пример типовых тестовых заданий:

№ п/п	Вопрос	Варианты ответа
1	Какой из перечисленных принципов НЕ относится к функционализму?	а) Объяснение языковой формы через её функции б) Генеративистский подход (акцент на врождённых структурах) в) Принцип экономии г) Мотивация грамматики дискурсом
2	Какое из утверждений верно характеризует проблему грамматикализации?	а) Это процесс превращения лексических единиц в грамматические

		маркеры. б) Это изучение только синхронных изменений в языке. в) Это анализ исключительно фонетических закономерностей. г) Это теория, отрицающая влияние дискурса на грамматику.
3	Какие три метафункции языка выделяются в СФЛ?	а) Фонетическая, морфологическая, синтаксическая б) Понятийная, межличностная, текстуальная в) Денотативная, коннотативная, экспрессивная г) Референциальная, эмоциональная, поэтическая
4	Какой элемент НЕ входит в кардиффскую модель функциональной грамматики?	а) Лексико-грамматика б) Грамматикализация в) Генеративные правила трансформации г) Классы и функции
5	Что такое фрейм в семантике Ч. Филлмора?	а) Морфологическая схема словообразования б) Когнитивная структура, организующая знания о типичной ситуации в) Синтаксическое дерево зависимостей г) Статистическая модель распределения слов в корпусе
6	Какой проект использует фреймовый подход для анализа лексики?	а) WordNet б) FrameNet в) BERT г) GPT
7	Что означает «профиль-база» в когнитивной грамматике?	а) Соотношение активной зоны значения и его когнитивной основы б) Частотность употребления слова в корпусе в) Морфологическую парадигму слова г) Синтаксическую функцию слова в предложении
8	Какая модель НЕ относится к грамматике конструкций?	а) Модель Филлмора б) Модель Хомского в) Модель Лакоффа-Гольдберга г) Модель У. Крофта
9	Что изучает коллострукционный анализ?	а) Распределение фонем в тексте б) Закономерности совместной встречаемости слов и грамматических структур в) Частотность отдельных морфем г) Диалектные различия в произношении

10	Какой инструмент НЕ используется для корпусного анализа?	а) AntConc б) Sketch Engine в) ELAN г) Word2Vec
----	--	--

Примеры типовых практических заданий:

1) Произведите корпусной анализ с целью выявления примеров проявления экономии и иконичности в языке:

- Найдите примеры проявления экономии в текстах XIX и XX веков: посчитайте частотность сокращений и использования клише, сравните длины слов и предложений в текстах, посчитайте статистики использования эллипсиса.

- Найдите примеры проявления иконичности в текстах XIX и XX веков: найдите образцы использования лексических и морфологических повторов, опишите закономерности их использования; проанализируйте соответствие использования порядка слова при передаче последовательности передаваемых событий; найдите примеры использования звукоподобия в текстах.

2) Сравните результаты в текстах XIX и XX века, опишите выявленные закономерности и ключевые различия в проанализированных текстах.

3.5. Рабочая программа дисциплины

«SQL и NoSQL базы данных»

Рабочая программа модуля «SQL и NoSQL базы данных» нацелена на формирование у обучающихся знаний и навыков проектирования и использования SQL и NoSQL баз данных, построения запросов, работы с данными и интеграции СУБД с клиентскими приложениями.

Тема 1 «Модели данных. Файловые системы и базы данных»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
1.1.	Понятие модели данных, виды (иерархическая, сетевая, реляционная, документоориентированная); отличие файловых систем от СУБД, преимущества использования СУБД.	2	ЭО (2 часа)
Перечень лабораторных работ			
1.1.	Создание и настройка БД. Установка и настройка PostgreSQL/MySQL; подключение к серверу; создание базы данных, схем и таблиц; определение типов данных; наполнение тестовыми данными; выполнение базовых CRUD-операций.	4	ЭО (4 часа)
Тема 2 «Модель ANSI/SPARC. СУБД»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
2.1.	Уровни модели: внешний, концептуальный, внутренний; архитектура и классификация СУБД; примеры популярных систем (PostgreSQL, MySQL, MongoDB).	2	ЭО (2 часа)
Перечень лабораторных работ			
2.1.	Реляционная алгебра и базовые запросы SQL. Выполнение операций выборки, соединения, фильтрации, группировки; использование WHERE, ORDER BY, GROUP BY, HAVING; работа с агрегатными функциями.	4	ЭО (4 часа)
Тема 3 «Основы реляционной алгебры. SQL: синтаксис и основные		Кол-во	Образовательные

операторы»		часов	технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
3.1.	Операции выборки, проекции, соединения, объединения; структура SQL-запроса; операторы SELECT, INSERT, UPDATE, DELETE; агрегатные функции и группировка данных.	2	ЭО (2 часа)
Перечень лабораторных работ			
3.1.	Взаимодействие СУБД с клиентскими приложениями. Подключение БД из Python/Java; выполнение SQL-запросов из приложения; обработка результатов и вывод отчёта.	4	ЭО (4 часа)
Тема 4 «Сложные запросы и NoSQL»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
4.1.	Подзапросы, CTE, индексы и оптимизация запросов; обзор NoSQL (MongoDB), особенности документоориентированных БД, базовые CRUD-операции в MongoDB.	2	ЭО (2 часа)
Перечень лабораторных работ			
4.1.	Сложные запросы SQL. Подзапросы, CTE, оконные функции, использование индексов; анализ планов выполнения запросов (EXPLAIN).	4	ЭО (4 часа)
4.2.	Работа с NoSQL БД. Создание коллекций в MongoDB, вставка, выборка, обновление, удаление документов; фильтрация и сортировка данных; вложенные документы.	4	ЭО (4 часа)
Промежуточная аттестация		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
5.1.	Промежуточная аттестация в формате тестирования	2	ЭО (2 часа)

3.5.1. Учебно-методическое и информационное обеспечение учебной дисциплины

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

Перечень учебной литературы:

- 1) Кузнецов, С. Е. Язык SQL. Полное руководство / С. Е. Кузнецов. — 4-е изд., испр. и доп. — Санкт-Петербург: Питер, 2023. — 720 с. — (Серия «Программист профессионал»).
- 2) Миклис, С. В. PostgreSQL 15: руководство администратора и разработчика / С. В. Миклис. — Москва: ДМК Пресс, 2023. — 496 с.
- 3) Чекмарев, А. В. NoSQL для разработчиков: MongoDB, Redis, Cassandra / А. В. Чекмарев. — Москва: БХВ-Петербург, 2022. — 384 с.
- 4) Elmasri, R. Fundamentals of Database Systems / R. Elmasri, S. B. Navathe. — 8th ed. — Boston: Pearson, 2023. — 1200 p.
- 5) Coronel, C. Database Systems: Design, Implementation, and Management / C. Coronel, S. Morris. — 14th ed. — Boston: Cengage, 2022. — 880 p.
- 6) Date, C. J. Database Design and Relational Theory: Normal Forms and All That Jazz / C. J. Date. — 3rd ed. — Beijing: O'Reilly Media, 2021. — 600 p.

3.5.2. Материально-техническое обеспечение учебной дисциплины

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE). Предусмотрен формат BYOD.

3.5.3. Система контроля и оценивания

Оценка качества освоения дисциплины включает промежуточную и итоговую аттестацию обучающихся.

Промежуточный контроль знаний проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся.

Итоговая аттестация по дисциплине проводится в форме тестирования, расположенного на СДО iSpring по адресу <https://edu.miet.ru>. Для получения зачета по дисциплине необходимо верно ответить не менее, чем на 70% вопросов итогового тестирования.

Пример типовых тестовых заданий:

№ п/п	Вопрос	Варианты ответа
1	Какая операция реляционной алгебры используется для выборки определённых столбцов?	A) Join B) Select C) Project D) Union
2	Какой оператор SQL используется для сортировки результата?	A) SORT BY B) ORDER BY C) GROUP BY D) ARRANGE
3	Что означает уровень «внешний» в модели ANSI/SPARC?	A) Физическая реализация данных B) Представление данных для пользователя C) Логическая структура базы D) Оптимизация запросов
4	Какая СУБД является документоориентированной?	A) MySQL B) PostgreSQL C) MongoDB D) Oracle
5	Какой оператор используется для объединения результатов двух запросов?	A) UNION B) JOIN C) INTERSECT D) MERGE

Примеры типовых практических заданий:

- 1) Спроектируйте схему БД для онлайн-магазина, определите ключи.
- 2) Составьте SQL-запрос для получения списка клиентов, оформивших заказы в прошлом месяце.
- 3) Оптимизируйте запрос с использованием индекса.
- 4) Создайте коллекцию MongoDB и добавьте в неё 5 документов.
- 5) Напишите код на Python для подключения к СУБД и выполнения запроса.

3.6. Рабочая программа дисциплины

«Автоматическая обработка текста»

Рабочая программа модуля «Автоматическая обработка текста» нацелена на формировании знаний и умений у обучающихся по применению программных алгоритмов обработки данных для инженерно-научных расчетов, подбору и использованию специализированных библиотек для научных вычислений, а также разработки программных алгоритмов с использованием методов математической оптимизации.

Тема 1 «Введение в обработку текста и естественного языка»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
1.1.	Определение и области применения NLP (анализ текста, классификация, анализ комментариев). Текст как данные: структурированные и неструктурированные данные, кодировки (ASCII, Unicode). Этапы обработки текста: предобработка, токенизация, нормализация, морфология, синтаксис, семантика. Предобработка текста: очистка (удаление шума, URL, пунктуации), нормализация регистра, удаление стоп-слов. Введение в анализ текстов: комментарии, большие тексты, сетевой анализ. Python для обработки текста: библиотеки re, string, collections, numpy	2	ДОТ (2 часа)
Перечень лабораторных работ			
1.1.	Чтение/запись текстовых файлов в Python (open, encode, decode). Предобработка: удаление URL, email, чисел, специальных символов с помощью регулярных выражений (re). Токенизация: разделение текста на слова и предложения (правила). Удаление стоп-слов: создание собственного списка стоп-слов.	2	ЭО (2 часа)
Тема 2 «Распознавание именованных сущностей и метрики оценки»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
2.1.	Распознавание именованных сущностей (NER): правило-based и статистические подходы. Метрики оценки: точность, полнота, F1-мера, матрицы ошибок. Применение NER в анализе комментариев и больших текстов. Корпусная лингвистика: создание и анализ корпусов	2	ДОТ (2 часа)
Перечень лабораторных работ			
2.1.	Распознавание именованных сущностей (NER). Содержание: Правило-based NER: использование регулярных выражений и словарей для сущностей (люди, организации). Статистический подход: применение HMM для NER.	2	ЭО (2 часа)
Тема 3 «Морфологический анализ и неглубокие векторные представления»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
3.1.	Морфология в NLP: флексия, деривация, композиты. Лемматизация vs. стемминг: словарные подходы. Неглубокие векторные представления: «мешок слов» (bag-	2	ДОТ (2 часа)

	of-words), TF-IDF, их применение в анализе текстов. Конечные автоматы (FSA) и трансдьюсеры (FST) для морфологического анализа. Введение в анализ больших текстов: масштабируемость, обработка объемных данных.		
Перечень лабораторных работ			
3.1.	Анализ комментариев: особенности коротких текстов (шум, смайлики, неформальность). Продвинутое токенизация: обработка сокращений, смайлов, хэштегов. Стемминг: реализация правило-based стеммера (например, для удаления -ing, -ed). Лемматизация.	2	ЭО (2 часа)
3.2.	Реализация конечного автомата в Python для морфологических задач (например, распознавание глагольных форм). Правило-based POS-тегирование: создание теггера с использованием словарей и регулярных выражений.	2	ЭО (2 часа)
Тема 4 «Работа с API и статистические модели»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
4.1.	Работа с API: получение текстовых данных через API (например, новостные или социальные медиа API).	2	ДОТ (2 часа)
Перечень лабораторных работ			
4.1.	Основы веб-скрейпинга: использование библиотек requests и BeautifulSoup для сбора текстов. Парсинг HTML: извлечение текстов из веб-страниц (например, комментариев, статей). Этические аспекты скрейпинга: ограничения, robots.txt. Понимание структуры веб-данных.	2	ЭО (2 часа)
4.2.	Получение текстовых данных через API (например, используя requests для публичного API). Реализация HMM в Python для POS-тегирования или NER.	2	ЭО (2 часа)
4.3.	Классификация текста: бинарные и многоклассовые задачи (например, анализ тональности комментариев). Извлечение признаков: «мешок слов», TF-IDF, реализация в Python.	2	ЭО (2 часа)
4.4.	Реализация базового парсера для контекстно-свободных грамматик (например, СΥК). Извлечение коллокаций: точечная взаимная информация (PMI).	2	ЭО (2 часа)
Тема 5 «Синтаксический анализ и грамматика зависимостей»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
5.1.	Синтаксический анализ: контекстно-свободные грамматики (CFG), грамматика зависимостей. Алгоритмы парсинга: СΥК, Earley для CFG; переходного-ориентированный парсинг зависимостей. Грамматика зависимостей: представление структуры предложения в виде графов. Введение в сетевой анализ: графы для анализа текстовых связей. Алгоритмы парсинга: СΥК, Earley для CFG; переходного-ориентированный парсинг зависимостей. Грамматика зависимостей: представление структуры предложения в виде графов. Введение в сетевой анализ: графы для анализа текстовых связей.	2	ДОТ (2 часа)
Перечень лабораторных работ			
5.1.	Реализация упрощенной версии LDA (например, на основе	2	ЭО (2 часа)

	term co-occurrence). Частотный анализ: слова, фразы, конкордансы в большом корпусе.		
5.2.	Экстрактивная суммаризация: выбор ключевых предложений на основе весов. Алгоритмы ранжирования предложений: подсчет частот слов, позиционные метрики.	2	ЭО (2 часа)
Тема 6 «Тематическое моделирование и анализ больших текстов»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
6.1.	Тематическое моделирование: основы Latent Dirichlet Allocation (LDA), применение в анализе текстов. Анализ больших текстов: масштабируемость, работа с большими корпусами. Меры сходства текста: косинусное сходство, коэффициент Жаккара. Неглубокие векторные представления: TF-IDF для больших текстов.	2	ДОТ (2 часа)
Перечень лабораторных работ			
6.1.	Анализ тональности: классификация комментариев на позитивные, негативные, нейтральные. Использование словарей тональности и статистических методов.	2	ЭО (2 часа)
6.2.	Особенности обработки многоязычных текстов: кодировки, токенизация, морфология. Реализация токенизатора для текста на двух языках (например, английский и русский).	2	ЭО (2 часа)
6.3.	Извлечение ключевых слов: методы на основе TF-IDF, RAKE (Rapid Automatic Keyword Extraction). Реализация алгоритма RAKE для извлечения ключевых фраз.	2	ЭО (2 часа)
Тема 7 «Суммаризация, QA-системы и графы знаний»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
7.1.	Суммаризация текстов: экстрактивные и статистические методы. QA-системы: правило-based подходы к вопросно-ответным системам. Графы знаний: структура, создание, применение в NLP (например, для анализа связей в текстах). Перспективы NLP: гибридные подходы, мультязычная обработка.	2	ДОТ (2 часа)
Перечень лабораторных работ			
7.1.	Особенности текстов социальных медиа: неформальность, хэштеги, упоминания. Анализ связей в текстах: построение графа упоминаний или хэштегов.	2	ЭО (2 часа)
Промежуточная аттестация		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
8.1.	Промежуточная аттестация в формате тестирования	2	ЭО (2 часа)

3.6.1. Учебно-методическое и информационное обеспечение учебной дисциплины

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

Перечень учебной литературы:

- 1) Маннинг, К., Шютце, Х. Основы статистической обработки естественного языка: учебное пособие / пер. с англ. — Москва: ДМК Пресс, 2021. — 640 с.
- 2) Jurafsky, D., Martin, J. H. Speech and Language Processing. — 3rd ed. — London: Pearson, 2023. — 1024 p.
- 3) Васильева, А. Н., Писаревский, С. А. Введение в автоматическую обработку текста: учебное пособие. — Санкт-Петербург: Лань, 2020. — 224 с.
- 4) Bird, S., Klein, E., Loper, E. Natural Language Processing with Python. — 2nd ed. — Beijing: O'Reilly Media, 2021. — 504 p.
- 5) Ключев, Е. В., Гаврилова, Т. А. Технологии обработки естественного языка: монография. — Москва: Наука, 2022. — 312 с.
- 6) Aggarwal, C. C., Zhai, C. Mining Text Data. — New York: Springer, 2019. — 524 p.
- 7) Переверзев, М. А. Компьютерная лингвистика и обработка текстовой информации: учебное пособие. — Москва: Юрайт, 2023. — 310 с.

3.6.2. Материально-техническое обеспечение учебной дисциплины

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE). Предусмотрен формат BYOD.

3.6.3. Система контроля и оценивания

Оценка качества освоения дисциплины включает промежуточную и итоговую аттестацию обучающихся.

Промежуточный контроль знаний проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся.

Итоговая аттестация по дисциплине проводится в форме тестирования, расположенного на СДО iSpring по адресу <https://edu.miet.ru>. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Пример типовых тестовых заданий:

№ п/п	Вопрос	Варианты ответа
1	Что представляет собой токенизация в автоматической обработке текста?	a) Преобразование текста в аудиофайл b) Разделение текста на предложения и слова c) Исправление орфографических ошибок d) Замена слов синонимами
2	Какой алгоритм чаще всего используется для стемминга в английском языке?	a) Porter Stemmer b) Levenshtein Distance c) K-means clustering d) LDA
3	Что такое стоп-слова?	a) Слова с наибольшей частотой в тексте

		b) Слова, которые невозможно токенизировать c) Специальные слова для кодирования текста d) Технические термины
4	Как называется процесс автоматического определения части речи слова в тексте?	a) POS-tagging b) Lemmatization c) Parsing d) Tokenization
5	Что делает алгоритм TF-IDF?	a) Строит синтаксическое дерево b) Определяет частоту слова в документе с учётом важности в корпусе c) Объединяет слова в биграммы d) Переводит текст на другой язык

Примеры типовых практических заданий:

- 1) Выполните токенизацию предложенного текста, удалите стоп-слова и представьте результат в виде списка токенов.
- 2) Постройте частотный словарь для текста и выделите 10 наиболее часто встречающихся слов.
- 3) Выполните лемматизацию текста с использованием выбранной библиотеки Python (например, spaCy или rymorphy2).
- 4) Рассчитайте TF-IDF для набора предложений и определите, какие слова являются наиболее информативными.
- 5) Разработайте простой алгоритм для выделения именованных сущностей (Named Entity Recognition) в предложенном тексте.

3.7. Рабочая программа дисциплины

«Нейронные сети в обработке естественного языка»

Рабочая программа модуля «Нейронные сети в обработке естественного языка» направлена на получение студентами знаний и умений в области обработки естественного языка с использованием современных нейронных сетей, включая большие языковые модели, методы их сжатия, интеграции и развертывания.

Тема 1 «Введение в обработку текста и естественного языка»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
1.1.	История развития сферы обработки естественного языка. Основные задачи области обработки естественного языка. Сферы применения инструментов обработки естественного языка.	2	ДОТ (2 часа)
Перечень лабораторных работ			
1.1.	Терминологический аппарат инженера в области обработки естественного языка. Составление глоссария. Токенизация. Стемминг. Лемматизация. Стоп-слова. N-граммы. Корпус. Датасет. Парсинг. Разметка. Препроцессинг.	2	ЭО (2 часа)
1.2.	Освоение работы с основными библиотеками и	2	ЭО (2 часа)

	программными средствами для лингвистического анализа текста (Mystem, Py morphology, NLTK, SpaCy, Natasha).		
Тема 2 «Векторные представления текста»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
2.1.	Необходимость построения векторного представления текста в компьютерной обработке естественного языка. Оценка семантической близости слов. Базовые методы построения векторных представлений (Bag-of-Words, One-Hot Encoding, TF-IDF). Статистические и контекстные методы построения векторных представлений. Статистические модели (Word2Vec, FastText, GloVe). Контекстные эмбединги (ELMo, BERT).	2	ДОТ (2 часа)
Перечень лабораторных работ			
2.1.	Построение векторных представлений для заданного текста с помощью статических моделей. Настройка параметров. Вычисление семантически близких слов на основе векторного сходства. Применение методов снижения размерности матриц для визуализации результатов.	2	ЭО (2 часа)
2.2.	Построение векторных представлений для заданного текста с помощью контекстных моделей. Настройка параметров, выявление семантически близких слов, визуализация результатов.	2	ЭО (2 часа)
Тема 3 «Сверточные нейронные сети в обработке естественного языка»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
3.1.	Классификация как одна из базовых задач обработки текста. Анализ тональности текста, определение авторства, фильтрация спама, морфологический анализ текста. Классические подходы к решению задачи классификации. Введение в работу сверточных нейросетей.	2	ДОТ (2 часа)
Перечень лабораторных работ			
3.1.	Разбор работы сверточных нейросетей на примере обработки изображений.	2	ЭО (2 часа)
3.2.	Реализация алгоритма анализа тональности текстов отзывов на товары с помощью сверточных нейросетей. Метрики оценки качества классификации текстов.	2	ЭО (2 часа)
Тема 4 «Рекуррентные нейронные сети в обработке естественного языка»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
4.1.	Моделирование языка и генерация текста: сферы применения и вызовы. Основы теории вероятностей для построения моделей для предсказания следующего слова. Рекуррентная нейросеть. Проблема взрывающегося и исчезающего градиента. Проблема «бутылочного горлышка». Долгая краткосрочная память. Управляемый рекуррентный блок.	2	ДОТ (2 часа)
Перечень лабораторных работ			

4.1.	N-граммные языковые модели.	2	ЭО (2 часа)
4.2.	Применение рекуррентных нейросетей в машинном переводе и в разработке диалоговых систем.	2	ЭО (2 часа)
Тема 5 «Механизм внимания и архитектура «Трансформер»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
5.1.	Архитектура Sequence-to-Sequence. Энкодеры и декодеры в машинном переводе. Механизм внимания: история возникновения, основные принципы работы, применение в нейросетевом машинном переводе. Архитектура «Трансформер» для машинного перевода. Основные компоненты архитектуры: внутреннее внимание, многоголовое внимание, блоки кодера и декодера, позиционное кодирование.	2	ДОТ (2 часа)
Перечень лабораторных работ			
5.1.	Исследование основных компонентов архитектуры «Трансформер» и используемых в ней вычислений. Запуск и оценка обученной модели; сравнение результатов работы модели со сверточными и рекуррентными нейросетями на задачах классификации, реферирования текста и машинного перевода.	2	ЭО (2 часа)
5.2.	Сравнение результатов работы модели со сверточными и рекуррентными нейросетями на задачах классификации, реферирования текста и машинного перевода.	2	ЭО (2 часа)
Тема 6 «Трансферное обучение»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
6.1.	Проблемы недостаточности данных в машинном обучении на основе признаков и в обработке естественного языка. Машинное обучение с возможностью «переноса» знаний. История развития трансферного обучения. Разновидности трансферного обучения. Преимущества использования архитектуры «Трансформер» для трансфера знаний. Обучение моделей с нуля и тонкая настройка нейросетей. Кодеры, декодеры и автоэнкодеры.	2	ДОТ (2 часа)
Перечень лабораторных работ			
6.1.	Знакомство с платформой HuggingFace. Работа с токенизаторами. Выбор модели на хабе. Изучение карточки модели. Тонкая настройка и оценка модели с помощью инструментов HuggingFace.	2	ЭО (2 часа)
6.2.	Выгрузка и публикация набора данных на HuggingFace Datasets.	2	ЭО (2 часа)
Тема 7 «Большие языковые модели»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
7.1.	Генеративные модели, промпты, техники zero-, one- и few-shot learning. Обучение с подкреплением на основе отзывов людей. Вызовы при использовании и интеграции больших языковых моделей. Гиперпараметры больших языковых	2	ДОТ (2 часа)

	моделей. Архитектура Mixture of Experts. Механизм Reasoning. Методы сжатия больших языковых моделей (квантование, дистилляция и проч.). Методы дообучения больших языковых моделей (LoRA, QLoRA, адаптеры).		
Перечень лабораторных работ			
7.1.	Работа с API для получения доступа к большим языковым моделям.	2	ЭО (2 часа)
7.2.	Разработка простого Telegram-бота на основе больших языковых моделей.	2	ЭО (2 часа)
Тема 8 «Информационная и кибербезопасность в NLP. Перспективы и вызовы использования больших языковых моделей»		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
Содержание лекционных занятий			
8.1.	Retrieval-Augmented Generation. Механизм извлечения информации и генерации текста на основе внешнего источника данных. Методы сбора и хранения данных для систем данного типа. Мультиагентная система. Разработка сервиса на основе систем искусственного интеллекта. Гардрейлы. Методы обеспечения информационной и кибербезопасности систем на основе больших языковых моделей.	2	ДОТ (2 часа)
Перечень лабораторных работ			
8.1.	Разработка пайплайна системы Retrieval-Augmented Generation.	2	ЭО (2 часа)
8.2.	Разработка прототипа сервиса на основе больших языковых моделей.	2	ЭО (2 часа)
Промежуточная аттестация		Кол-во часов	Образовательные технологии, в том числе ЭО и (или) ДОТ с указанием часов
9.1.	Промежуточная аттестация в формате тестирования	2	ЭО (2 часа)

3.7.1. Учебно-методическое и информационное обеспечение учебной дисциплины

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

Перечень учебной литературы:

- 1) Jurafsky D., Martin J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. URL: <https://web.stanford.edu/~jurafsky/slp3/> (дата обращения 28.07.2025)
- 2) Schütze H., Manning C. D., Raghavan P. Introduction to information retrieval. – Cambridge: Cambridge University Press, 2008. – Т. 39. – С. 234-265.
- 3) Goodfellow I. et al. Deep learning. – Cambridge: MIT press, 2016. – Т. 1. – №. 2.
- 4) Gao J., Galley M., Li L. Neural approaches to conversational AI //The 41st international ACM SIGIR conference on research & development in information retrieval. – 2018. – С. 1371-1374.

Информационные ресурсы:

- 1) Образовательный курс на платформе “DeepLearning.AI”. Natural Language Processing Specialization. URL: <https://www.deeplearning.ai/courses/natural-language-processing-specialization/> (дата обращения: 28.07.2025).

3.7.2. Материально-техническое обеспечение учебной дисциплины

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE). Предусмотрен формат BYOD.

3.7.3. Система контроля и оценивания

Оценка качества освоения дисциплины включает промежуточную и итоговую аттестацию обучающихся.

Промежуточный контроль знаний проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся.

Итоговая аттестация по дисциплине проводится в форме тестирования, расположенного на СДО iSpring по адресу <https://edu.miet.ru>. Для получения зачета по дисциплине необходимо верно ответить не менее, чем на 70% вопросов итогового тестирования.

Пример типовых тестовых заданий:

№ п/п	Вопрос	Варианты ответа
1	Какой из методов НЕ используется для построения эмбедингов?	1) Word2Vec 2) TF-IDF 3) Random Forest 4) GloVe
2	Какой механизм лежит в основе архитектуры «Трансформер»?	1) Сверточные слои 2) Self-Attention 3) Рекуррентные связи 4) Полносвязные сети
3	Какой формат используется для квантованных языковых моделей в llama.cpp?	1) ONNX 2) GGUF 3) HDF5 4) JSON

Примеры типовых практических заданий:

Сравните методы токенизации:

1) Примените токенизацию с помощью NLTK, SpaCy и BERT-токенизатора к одному и тому же тексту.

2) Сравните результаты, выделите преимущества и недостатки каждого метода.

Данное практическое задание предполагает выполнение следующих этапов:

1) Установка библиотек и настройка окружения.

2) Выполнение токенизации с помощью NLTK.

3) Выполнение токенизации с помощью SpaCy.

4) Составление сравнительной таблицы, выделение таких критериев, как точность токенизации, скорость выполнения функции, анализ контекста, простота использования, сферы применения.

4. УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

Описание учебно-методического и материально-технического обеспечения программы приведено в рабочих программах учебных дисциплин (модулей), практик/ стажировок.

5. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

5.1. Фонды оценочных средств для итоговой аттестации

5.1.1. Практические задания

Итоговая аттестация проводится в форме демонстрационного экзамена в формате комплексного практического задания (кейса) или хакатона и презентационной защиты.

Демонстрационный экзамен моделирует выполнение проекта в условиях, приближенных к профессиональной деятельности. Участники работают в командах (3–5 человек). Результатом является готовый программный продукт/прототип, сопровождаемый документацией и устной защитой.

Примеры заданий:

1) На основании предоставленного набора данных (тексты новостных статей и метаданные) спроектировать и реализовать систему автоматической классификации текстов по тематикам.

В рамках выполнения задания необходимо:

- Предварительно обработать данные (токенизация, лемматизация, удаление стоп-слов).
- Сохранить предобработанные данные в реляционную БД (PostgreSQL) с корректной структурой и индексами.
- Разработать Python-скрипт, который выполняет анализ текста и классифицирует его по заданным категориям.

- Для хранения метаданных эксперимента использовать NoSQL БД (MongoDB).

- Подготовить визуализацию результатов и итоговую презентацию.

2) На основании предоставленного корпуса отзывов о товарах/услугах на русском языке разработать систему автоматического анализа тональности (позитив, негатив, нейтрально) с использованием инструментов обработки текста, баз данных и визуализации.

В рамках выполнения задания требуется:

- Предобработка текста. Очистка текста от лишних символов и HTML-тегов. Токенизация, лемматизация, удаление стоп-слов.
- Хранение данных. Проектирование структуры реляционной БД для хранения исходных и обработанных данных. Загрузка данных в PostgreSQL.
- Разработка анализа тональности. Использование Python-библиотек (например, NLTK, spaCy, DeepPavlov) для классификации отзывов. Определение тональности с учётом ключевых слов и/или модели машинного обучения.
- NoSQL-хранилище. Сохранение статистики по тональности и метаданных эксперимента в MongoDB.

- Визуализация и защита. Построение диаграмм (доля позитивных/негативных отзывов). Презентация алгоритма, результатов и выводов комиссии.

3) Разработать прототип поисковой системы, позволяющей осуществлять тематический поиск по базе научных статей с ранжированием результатов по релевантности.

В рамках выполнения задания требуется:

- Загрузка и подготовка данных. Получение корпуса статей в формате JSON/CSV. Нормализация текста, извлечение ключевых слов и фраз.
- Проектирование базы данных. Создание структуры реляционной БД (PostgreSQL) для хранения метаданных (название, автор, ключевые слова, аннотация). Хранение полнотекстового индекса.
- Разработка поискового алгоритма. Реализация полнотекстового поиска средствами PostgreSQL (tsvector, tsquery) или Elasticsearch (по желанию). Сортировка результатов по релевантности.
- NoSQL-хранилище. Сохранение истории поисковых запросов пользователей в MongoDB.

- Интерфейс и защита. Создание консольного или веб-интерфейса для ввода запроса и вывода результатов. Подготовка презентации к защите.

4) Автоматическая генерация резюме текста научных статей

В рамках выполнения задания требуется:

- Загрузка и предобработка данных: получить корпус статей (в формате PDF/JSON/CSV), выделить заголовок, аннотацию и основной текст. Очистить текст от лишних символов, провести токенизацию, лемматизацию, удаление стоп-слов.

- Хранение данных: спроектировать структуру реляционной БД (PostgreSQL) для хранения исходных и обработанных данных, включая текст статей и сгенерированные краткие резюме.

- Разработка алгоритма: применить Python-библиотеки для генерации кратких резюме (например, gensim, HuggingFace Transformers). Реализовать как экстрактивное, так и абстрактное суммирование.

- NoSQL-хранилище: сохранять метрики качества (ROUGE, BLEU) и историю экспериментов в MongoDB.

- Визуализация и защита: подготовить таблицы и графики с сравнением качества разных алгоритмов, представить результаты в презентации.

5) Автоматическое выявление фейковых новостей

В рамках выполнения задания требуется:

- Подготовка корпуса данных: загрузить набор новостных статей (CSV/JSON), помеченных как достоверные и фейковые. Провести очистку текста (удаление HTML, спецсимволов), токенизацию и лемматизацию.

- Хранение данных: разработать схему PostgreSQL для хранения исходных текстов, признаков и результатов классификации.

- Модель машинного обучения: реализовать классификатор (Logistic Regression, SVM, BERT) для автоматического определения фейковых новостей.

- NoSQL-хранилище: сохранять результаты классификации и логи экспериментов в MongoDB.

- Визуализация и защита: построить графики точности (Precision, Recall, F1-score), разработать дашборд с результатами, подготовить презентацию для защиты.

5.2. Критерии оценивания

Критерии оценивания итоговой аттестации:

№	Критерий	Показатели результата	Вес критерия (макс. балл)	Высокий уровень (8-10б.)	Базовый уровень (5-7б.)	Низкий уровень (1-4б.)
1	Анализ данных и постановка задачи	Определяет цели, формулирует требования, выбирает методы обработки	15	Цели и задачи чёткие, методы обоснованы	Цели и задачи определены, методы частично обоснованы	Цели нечеткие, методы случайны
2	Реализация алгоритмов	Пишет рабочий и оптимизированный код, реализует алгоритмы	15	Код рабочий, оптимизирован, документирован	Код рабочий, но есть избыточность	Код с ошибками, неполный функционал
3	Применение языков программирования	Использует Python и другие инструменты для решения задач	15	Языки применяются эффективно, библиотеки используются по назначению	Применение с частичными ошибками	Конструкции неверны, результата нет
4	Работа с базами данных (SQL/NoSQL)	Создаёт структуру БД, формирует корректные запросы	15	Структура корректна, запросы оптимальны	Структура верна, но неоптимальна	Структура и запросы некорректны

5	Применение ИИ и машинного обучения	Использует готовые модели или разрабатывает собственные	15	Модель выбрана/разработана верно, результаты воспроизводимы	Модель верна, но результаты неполные	Модель неверна или не реализована
6	Качество анализа и интерпретации	Применяет корректные методы, делает выводы	15	Методы и выводы верны	Методы верны, выводы частично обоснованы	Методы неверны, выводы отсутствуют
7	Презентация проекта	Представляет результаты, аргументирует решения	10	Презентация логична, ответы уверенные	Презентация понятна, ответы частично верны	Презентация бессистемна, ответы отсутствуют

Максимальный балл за итоговую аттестацию – 100. Менее 50 баллов соответствует оценке «неудовлетворительно», 51-69 баллов соответствуют оценке «удовлетворительно», 70-85 баллов оценке «хорошо», 86-100 баллов оценке «отлично».

Слушатели, получающие высшее образование впервые и(или) среднее профессиональное образование (за исключением лиц, имеющих среднее профессиональное и(или) высшее образование) и отчисленные из образовательной организации (исключением является академический отпуск) не могут продолжать обучение по дополнительной профессиональной программе – программе профессиональной переподготовки.

Лицам, не прошедшим итоговую аттестацию или получившим на итоговой аттестации неудовлетворительные результаты, а также лицам, освоившим часть программы, выдается справка об обучении или о периоде обучения.

Лицам, завершившим обучение по программе, и успешно прошедшим итоговую аттестацию, присваивается дополнительная квалификация, установленная программой.

При освоении программы параллельно с получением высшего образования диплом о профессиональной переподготовке выдается одновременно с получением соответствующего документа об образовании и о квалификации (за исключением лиц, имеющих среднее профессиональное или высшее образование).

Разработчики программы:

Ст. преподаватель Института ЛПО

К.Л. Грабарник

Ст. преподаватель кафедры ВМ №1

К.Н. Коночкин

Согласовано:

Директор ИДСВО

И.В. Дарда

Руководитель центра «Цифровая кафедра»

В.В. Кокин

Директор Института ЛПО

М.Г. Евдокимова